

WHAT IS CLAIMED IS:

1. A server load balancing apparatus using Multi-Protocol Label Switching (MPLS) session labels, the server load
5 balancing apparatus being located in front of a plurality of servers connected to a plurality of clients through the Internet, analyzing requests from the clients and distributing the requests among the servers, comprising:

a packet analyzing unit for inspecting whether a session
10 label has been attached to a received packet, analyzing header information of the received packet to learn session information, and attaching a session label to a header of the received packet;

a load balancing processing unit for assigning a server
15 to a session of the received packet without the session label attached in view of load balancing according to a result of the inspection of the packet analyzing unit;

a session label switching unit for hardware-switching
the received packet with the session label attached using only
20 the session label information without a packet analysis process or server assignment process according to a result of the inspection of the packet analyzing unit;

a session managing unit for managing and maintaining
relevant information and states of sessions requested by the
25 clients; and

a session label managing unit for assigning the session
label, and withdrawing and managing session label not in use.

2. The server load balancing apparatus according to claim 1, wherein the packet analyzing unit comprises:

a session label inspecting unit for inspecting whether the session label has been attached to the received packet, transmitting the received packet to the session label switching unit to switch the received packet if the session label has been attached to the received packet, and transmitting the received packet to a packet contents analyzing unit if the session label is not attached to the received packet;

the packet contents analyzing unit for learning the session information by analyzing the header information of the received packet ranging from third to seventh layers of the received packet, inspecting whether a session of the received packet is a new session, transmitting the received packet of the new session to the load balancing processing unit to assign the server to the new session if the session is new, and transmitting the packet of an existing session to a predetermined server if the session is not new; and

a session label attaching unit for attaching the assigned session label to the header of the received packet.

3. The server load balancing apparatus according to claim 1, wherein the session label is an MPLS-based session label.

4. The server load balancing apparatus according to claim 1, wherein the load balancing processing unit comprises:

a load balancing algorithm unit for determining a load

balancing server using a specific algorithm in view of information including a round robin method, a minimally connected server, weights and response time from the server;

a server configuration/state managing unit for managing
5 configurations and states of the servers by performing real time server state monitoring or configuration management; and

a service acceptance control unit for refusing a service request of the new session if the existing session is serviced.

10

5. The server load balancing apparatus according to claim 1, wherein the session label switching unit performs label switching with reference to a value of the session label attached to the header of the received packet, and a label
15 switching table including information of line cards and ports through which the received packet is input/output.

6. The server load balancing apparatus according to claim 1, wherein the session managing unit recognizes start,
20 determination and interruption of the session, and adds, deletes and changes relevant information in the session table.

7. The server load balancing apparatus according to claim 1, wherein the assignment of the session label is performed in
25 such a way that a Client-To-Server (C2S) session label is assigned an odd number and a Server-To-Client (S2C) session label is assigned an even number obtained by adding 1 to the value of the C2S session label.

8. A server load balancing method using MPLS session labels, in which a plurality of clients are connected to a plurality of servers, and a server load balancing apparatus located in front of the plurality of the servers analyzes requests from the clients and distributes loads among the servers, comprising:

a first step of the server load balancing apparatus recognizing a new session by analyzing a header of a received packet and assigning a C2S session label when the client requests service from the server through the server load balancing apparatus;

a second step of the server load balancing apparatus assigning a specific server for servicing the session in view of load balancing, attaching the assigned C2S session label to the received packet, and transmitting the received packet with the C2S session label attached to the server;

a third step of the server having received the packet with the C2S session attached automatically assigning an S2C session label, that is, an opposite direction session label, with reference to a value of a session label of the received packet;

a forth step of the server processing the service request from the client, attaching the assigned S2C session label to the packet according to a result of the processing, and transmitting the received packet with the S2C session label to the server load balancing apparatus;

a fifth step of the server load balancing apparatus having received the packet with the S2C session label attached

from the server label switching the received packet to the client using the value of the session label;

a sixth step of the client having received the received packet with the S2C session label attached automatically
5 assigning the C2S session label, that is, an opposite direction session label to the received packet, with reference to the value of the session label of the received packet;

a seventh step of the client attaching the packet with the assigned C2S session label attached and transmitting the
10 packet with the assigned C2S session label to the server load balancing apparatus when the client transmits the packet to a destination server; and

an eighth step of the server load balancing apparatus having received the packet with C2S session label attached
15 from the client label switching the packet with C2S session label attached to the destination server;

wherein the server load balancing apparatus determines the server for connection using information of the session label with respect to the packet with the session label
20 attached.

9. The server load balancing method according to claim 8, wherein it is inspected whether the MPLS session label has been attached to the packet input into the server load
25 balancing apparatus, and the packet with the MPLS session label attached is fast-switched using only information of the session label.

10. The server load balancing method according to claim 8, wherein it is inspected whether the MPLS session label has been attached to the packet input into the server load balancing apparatus, and only the header of the packet header without the session label attached is selectively analyzed.

11. The server load balancing method according to claim 8, wherein the step of assigning the specific server comprises the step of determining whether to accept or refuse the session of only the packet without the session label attached.

12. The server load balancing method according to claim 8, wherein the server load balancing apparatus omits a function of performing mapping between a virtual IP address and an IP addresses of the server in such a way that the server attaches the virtual IP address to the header of the packet with the session label attached.

13. The server load balancing method according to claim 8, wherein the C2S session label is assigned an odd number, and the S2C session label is automatically assigned a value obtained by adding 1 to the value of the C2S session label.

14. The server load balancing method according to claim 8 or 13, wherein the assignment of the bi-directional session labels (S2S and S2C) is performed by automatically recognizing the value of the opposite directional label without using an additional protocol for assigning a session label to a packet

in such a way the server and the client add 1 to and subtract 1 from the value of the session label that is attached to the packet received from an opposite party, respectively.

- 5 15. The server load balancing method according to claim 8, wherein the session label is attached to the header of the received packet according to a MPLS header configuration.